Approximate ML Decision Feedback Block Equalizer for Doubly Selective Fading Channels

Lingyang Song, Rodrigo C. de Lamare, Are Hjørungnes, and Alister G. Burr

Abstract

In order to effetively suppress intersymbol interference (ISI) at low complexity, we propose in this paper an approximate maximum likelihood (ML) decision feedback block equalizer (A-ML-DFBE) for doubly selective (frequency-selective, time-selective) fading channels. The proposed equalizer design makes efficient use of the special time-domain representation of the multipath channels through a matched filter, a sliding window, a Gaussian approximation, and a decision feedback. The A-ML-DFBE has the following features: 1) It achieves performance close to maximum likelihood sequence estimation (MLSE), and significantly outperforms the minimum mean square error (MMSE) based detectors; 2) It has substantially lower complexity than the conventional equalizers; 3) It easily realizes the complexity and performance tradeoff by adjusting the length of the sliding window; 4) It has a simple and fixed-length feedback filter. The symbol error rate (SER) is derived to characterize the behaviour of the A-ML-DFBE, and it can also be used to find the key parameters of the proposed equalizer. In addition, we further prove that the A-ML-DFBE obtains full multipath diversity.

Index Terms

Doubly selective fading channels, equalization, matched filter, linear MMSE, MMSE-DFE, maximum likelihood sequence estimation.

I. INTRODUCTION

Wireless communications often suffer from severe inter-symbol interference (ISI) due to doubly selective fading. In order to suppress the channel distortion, channel equalization techniques are

Lingyang Song is with School of Electrical Engineering and Computer Science, Peking University, Beijing, China (e-mail: lingyang.song@pku.edu.cn).

Are Hjøungnes is with UniK-University Graduate Center, University of Oslo, Norway (e-mail: arehj@unik.no).

Rodrigo C. de Lamare and Alister G. Burr are with the Department of Electronics, University of York, York, UK (e-mail: rcdl500@ohm.york.ac.uk; alister@ohm.york.ac.uk).

essential, and indeed have received considerable attention for many years. Maximum a priori probability (MAP) equalization is the optimum equalization procedure in terms of minimum symbol error rate (SER) [1], but requires a prohibitive computational complexity for many applications, being exponential in the channel length and constellation size. Maximum likelihood sequence estimation (MLSE) can obtain SER performance very close to MAP, but its complexity is still extremely high [2]. As a result, many sub-optimal, low-complexity equalization techniques have been proposed, such as the popular minimum mean square error decision-feedback equalizer, which is very effective in certain multipath environments and has a complexity that is only dependent on forward and backward filter lengths [3]. However, there is a non-negligible performance loss of MMSE based equalizers in comparison to MLSE [4] [5].

Further still, while lots of research have been conducted on the time-domain equalization, few works take the special form of the channel representation into good account. Two properties of the channel matrix in time domain are effectively utilized in this paper: 1) The Toeplitz-like channel matrix significantly contributes to the equalizer design; 2) The large number of zero elements reduces the computational complexity. As a result, we propose a robust *approximate ML based decision feedback block equalizer* (A-ML-DFBE) to combat ISI over doubly selective fading channels with low computational complexity. The proposed equalizer exploits substantial benefit from the special time domain representation of the multipath channels by using a *matched filter*, a *sliding window*, a *Gaussian approximation*, and a *decision feedback*. The main ideas are firstly to subtract the effect of the already-detected signals obtained from past decisions. This can be treated as a decision feedback process. Secondly we apply Gaussian approximation [6]–[9] to realize near maximum likelihood detection. The accuracy of this procedure can be improved by adjusting the length of the sliding window due to the central limit theorem. Consequently, a complexity and performance trade-off can be realized, and a convergence in SER performance can also be obtained by adjusting the length of the sliding window.

Note that [6] and [7] can be used only for frequency flat fading channels, and [8] aims to recover signals for multiuser systems. Although in [9] a probabilistic data association (PDA) based equalizer is reported, there are several major differences compared to the proposed approach: In [9], it requires to update the mean and the variance for all detected symbols; many iterations have to be used in order to make the performance converge; there is no feedback process; and no matched filter is employed. In [10], bidirectional arbitrated decision-feedback equalization (BAD) algorithm was presented which has complexity at least two times of the MMSE-DFE but can achieve better performance. In [11], a class of block DFE is presented for frequency domain equalization, but it assumes that the length of the channel, forward filter, and backward filter are infinitely long which is not practical. Besides, it requires large number of iterations to make the performance converge, which increases the system

delay and the computational complexity.

The rest of the paper is organized as follows: In Section II, we present the channel and signal models. The proposed A-ML-DFBE scheme and complexity comparisons are discussed in Section III. The performance is analyzed in Section IV. Simulation results are presented in Section V. In Section VI, we draw the main conclusions. The proof is given in the appendix.

Notation: Boldface upper-case letters denote matrices, boldface lower-case letters denote vectors, $\mathbb{C}^{i\times j}$ and $\mathbb{R}^{i\times j}$ denote the set of $i\times j$ complex and real matrices, respectively, $(\cdot)^T$ stands for transpose, $(\cdot)^*$ denotes complex conjugate, $(\cdot)^H$ represents conjugate transpose, $\mathbf{I}_i$ stands for an $i \times i$ identity matrix, $\mathbb{E}$ is used for expectation, $\mathtt{var}$ is used for variance, and $\|\mathbf{x}\|^2 = \mathbf{x}^H\mathbf{x}$.

II. Channel and Signal Models

The doubly selective fading channel can be modeled using a finite impulse response (FIR) filter

$$H(z,t) = \sum_{k=0}^{L-1} h_k(t) z^{-k}, \tag{1}$$

where $H(z,t)$ denotes the z transformation at time t, $h_i(t)$ represents the *i*-th path's channel coefficient, and the length of the FIR filter is L. For simplicity, we only consider a single input and single output system. The received signals can be written in vector form as (for convenience, we drop the time index for each transmission frame)

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{2}$$

where the received signals $\mathbf{r} = [r_1, \ldots, r_{N+L-1}]^T$, N is the length of $\mathbf{s}$, transmitted signals $\mathbf{s} = [s_1, \ldots, s_N]^T$, and $\mathbf{n} = [n_1, \ldots, n_{N+L-1}]^T$ whose elements are independent samples of a zero-mean complex Gaussian random variable with variance $\sigma^2 = \mathbb{E}[|s_k|^2]/\mathtt{SNR}$, in which $\mathbb{E}[|s_k|^2]$ represents the average power of the transmitted symbols from constellation $\mathcal{A}$. In this paper, we set $\mathbb{E}[|s_k|^2] = 1$. The time domain representation of the doubly selective fading channel $\mathbf{H} \in \mathbb{C}^{(N+L-1)\times N}$, can be written as

$$\mathbf{H} = \begin{bmatrix} h_1(0) & 0 & 0 & \ldots & 0 \\ h_2(0) & h_1(1) & 0 & \ldots & \vdots \\ \vdots & h_2(1) & h_1(2) & \ldots & 0 \\ h_L(0) & \vdots & h_2(2) & \ldots & h_1(N-1) \\ 0 & h_L(1) & \vdots & \ldots & h_2(N-1) \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & \ldots & h_L(N-1) \end{bmatrix}.$$

Note that $\mathbf{H}$ has a structure similar to Toeplitz form, and some form of guard interval is necessary to avoid inter-block interference between the received signals [5]. The symbols in (2) can be recovered

by MLSE [1]. Alternatively, they can also be decoded in complex form using standard zero forcing (ZF) or MMSE approaches, linear or decision feedback equalization [3].

III. Description of the Proposed Method

A. Approximate Maximum Likelihood Decision Feedback Block Equalizer (A-ML-DFBE)

The proposed equalization algorithm can be summarized into three steps: 1) Forward process, which builds up the forward filter by a temporal sub matched filter; 2) Decision feedback process, which cancels the interference by a fixed length backward filter, and 3) Approximate ML process, which realizes the final signal detection by the aid of Gaussian approximation. The detailed description of each step is given below.

1) Forward Process: Supposing we start decoding s_k, a temporal sub matched filter (forward filter) is applied to (2)

$$\mathbf{H}_k^H \mathbf{r} = \mathbf{H}_k^H \mathbf{H}\mathbf{s} + \mathbf{H}_k^H \mathbf{n}, \tag{3}$$

where $\mathbf{H}_k$ denotes the matrix of size $N \times L_f$, which is made of the entries in $\mathbf{H}$, from the k-th column to the $(k + L_f - 1)$-th column and from the 1-st row to the N-th. L_f ($L \leq L_f \leq N$) is the length of the sliding window that must be equal or larger than L for smaller inter-symbol interference and larger diversity gain, and smaller than or equal to N. When $L_f = N$, the matched filter becomes $\mathbf{H}^H$. For simplicity, we may rewrite (3) as

$$\mathbf{y}_k = \mathbf{J}\mathbf{s} + \widetilde{\mathbf{n}}_k, \tag{4}$$

where $\mathbf{y}_k = \mathbf{H}_k^H \mathbf{r} \in \mathbb{C}^{L_f \times 1}$, $\mathbf{J} = \mathbf{H}_k^H \mathbf{H} \in \mathbb{C}^{L_f \times N}$, and $\widetilde{\mathbf{n}}_k = \mathbf{H}_k^H \mathbf{n} \in \mathbb{C}^{L_f \times 1}$. We call this process *horizonal slicing*, since it takes L_f rows of $\mathbf{H}$. $\mathbf{J}$ is given by

$$\mathbf{J} = \begin{bmatrix} \mathbf{h}_k^H \mathbf{h}_1 & \cdots & \mathbf{h}_k^H \mathbf{h}_N \\ \vdots & \ddots & \vdots \\ \mathbf{h}_{k+L_f-1}^H \mathbf{h}_1 & \cdots & \mathbf{h}_{k+L_f-1}^H \mathbf{h}_N \end{bmatrix}, \tag{5}$$

where $\mathbf{h}_i \in \mathbb{C}^{(N+L-1)\times 1}$ denotes the i-th column of matrix $\mathbf{H}$. The length of the forward filter has been defined as L_f in (3).

2) Decision Feedback Process: The function of this step is to suppress the effects of the detected terms.

In order to further decrease the complexity of (4), we can just consider a certain number of the transmitted symbols, and have

$$\mathbf{y}_k \approx \mathbf{J}_k \mathbf{s}_k + \widetilde{\mathbf{n}}_k, \tag{6}$$

where $\mathbf{J}_k \in \mathbb{C}^{L_f \times (k+L_f-1)}$ can be constructed by taking the first column to the $k+L_f-1$-th column of $\mathbf{J}$ in (5), and $\mathbf{s}_k = [s_1, \ldots, s_{k+L_f-1}]^T$. We call this process as *vertical slicing*, since it takes $k + L_f - 1$

columns of **J**. Moreover, (6) can be decomposed with respect to each transmitted symbol

$$\mathbf{y}_k \approx \sum_{i=1}^{k+L_f-1} \mathbf{j}_i s_i + \widetilde{\mathbf{n}}_k, \tag{7}$$

where $\mathbf{j}_i \in \mathbb{C}^{L_f \times 1}$ stands for the *i*-th column of the matrix $\mathbf{J}_k$, and s_i represents the *i*-th transmitted symbol. Note that (7) is equivalent to (4) when the *vertical slicing* includes all the symbols in **J**, $L_f = N+1-k$, which implies the length of L_f will have some effect on the system performance, and the effect of L_f will be discussed further in the performance analysis and simulation results sections. We can write (7) as

$$\mathbf{y}_k \approx \sum_{i=1}^{k-1} \mathbf{j}_i s_i + \mathbf{j}_k s_k + \sum_{i=k+1}^{k+L_f-1} \mathbf{j}_i s_i + \widetilde{\mathbf{n}}_k, \tag{8}$$

where $\sum_{i=1}^{k-1} \mathbf{j}_i s_i$ stands for the detected terms that can be rebuilt by the past decisions, $\mathbf{j}_k s_k$ is the current target, and $\sum_{i=k+1}^{k+L_f-1} \mathbf{j}_i s_i$ represents the undetected terms. The function of the *feedback process* is to reconstruct $\sum_{i=1}^{k-1} \mathbf{j}_i s_i$ for later interference cancellation. Therefore, it is important to decide the length of the backward filter, L_b. Based on the expressions of **H** and (5), we have $\mathbf{j}_1 = \mathbf{j}_2 = \cdots = \mathbf{j}_{k-L-2} = \mathbf{0}$, and thus, the *length* of the *backward filter* L_b can be fixed at $L-1, (L > 1)$ to reconstruct the effects of past decisions. (8) can be rewritten by simplifying the detected terms

$$\mathbf{y}_k \approx \sum_{i=k-L_b}^{k-1} \mathbf{j}_i s_i + \mathbf{j}_k s_k + \sum_{i=k+1}^{k+L_f-1} \mathbf{j}_i s_i + \widetilde{\mathbf{n}}_k, \tag{9}$$

where L_b equals $L-1$. As in (9), $\sum_{i=k-L_b}^{k-1} \mathbf{j}_i s_i$ can be reconstructed from past decisions, the following past decision cancellation process can be applied

$$\widetilde{\mathbf{y}}_k = \mathbf{y}_k - \sum_{i=k-L_b}^{k-1} \mathbf{j}_i s_i. \tag{10}$$

The above process is very similar to the decision feedback cancellation process, but unlike MMSE-DFE, we do not need to calculate the coefficients of the feedback filter, moreover, the length of L_b is fixed at $L-1$, which means that only $L-1$ past decisions need to be fed back, which is much less than what is typically required by MMSE-DFE.

3) Approximate ML: This step aims to achieve near optimal detection by applying the Gaussian approximation. We substitute (9) into (10) and get

$$\widetilde{\mathbf{y}}_k = \mathbf{j}_k s_k + \sum_{i=k+1}^{k+L_f-1} \mathbf{j}_i s_i + \widetilde{\mathbf{n}}_k. \tag{11}$$

In order to decode s_k with low computational complexity while maintaining the performance comparable to the ML decoder, we treat the undetected terms $\sum_{i=k+1}^{k+L_f-1} \mathbf{j}_i s_i$ and the noise vector $\widetilde{\mathbf{n}}_{k,k+L_f-1}$ together as a new complex-valued Gaussian vector with matching mean and covariance matrix, such that (11) can be expressed as

$$\widetilde{\mathbf{y}}_k = \mathbf{j}_k s_k + \boldsymbol{\eta}_k, \tag{12}$$

where $\boldsymbol{\eta}_k$ represents a vector with size $L_f \times 1$ of zero-mean complex-valued Gaussian random variables with covariance

$$\boldsymbol{\Lambda}_k = \mathbf{J}_{k+1}\mathbf{J}_{k+1}^H + \sigma^2 \mathbf{J}_k^{'}, \tag{13}$$

where $\mathbf{J}_{k+1}$ can be constructed by using the $(k+1)$-th column as the $(k+L_f-1)$-th column of $\mathbf{J}_k$ and $\mathbf{J}_k^{'}$ can be obtained by taking the k-th column to the $(k+L_f-1)$-th column of $\mathbf{J}_k$. According to the *central limit theorem*, the *accuracy* of the Gaussian assumption can be improved by increasing the length of the *forward filter* (sliding window), L_f.

As $\boldsymbol{\eta}_k$ has an approximate Gaussian distribution, the likelihood function $p(\widetilde{\mathbf{y}}_k|s_k)$ is given by

$$p(\widetilde{\mathbf{y}}_k|s_k) \propto \mathtt{exp}\left(-(\widetilde{\mathbf{y}}_k - \mathbf{j}_k s_k)^H \boldsymbol{\Lambda}_k^{-1} \ (\widetilde{\mathbf{y}}_k - \mathbf{j}_k s_k)\right). \tag{14}$$

Finally, s_k can be recovered by the following *ML* detector

$$s_k = \underset{s_k \in \mathcal{A}}{\mathtt{arg\ min}} \left((\widetilde{\mathbf{y}}_k - \mathbf{j}_k s_k)^H \boldsymbol{\Lambda}_k^{-1} (\widetilde{\mathbf{y}}_k - \mathbf{j}_k s_k)\right), \tag{15}$$

at $k = N - L_f + 1$, there are no more new received signals outside the sliding window. So, we can then simply decode each undetected symbol by treating the rest as Gaussian term and removing the effects of the detected symbols. This decoding process is very similar to the case of $k < N - L_f + 1$ by fixing the sliding window. The overall A-ML-DFBE algorithm is summarized in Table I.

B. Computational Complexity Analysis

Before we show the complexity comparisons, we present how to further reduce the proposed equalizer complexity. Note that, in (9), the detected terms $\sum_{i=k-L_b}^{k-1} \mathbf{j}_i s_i$ can be rewritten as $\mathbf{J}_{k-1}\mathbf{s}_{k-L_b}$ with size $L_f \times 1$, where $\mathbf{s}_{k-1} = [s_{k-L_b}, \ldots, s_{k-1}]^T$ has size $L_b \times 1$. With respect to the diagonal element $\mathbf{h}_g^H \mathbf{h}_g$ in $\mathbf{J}$, when $g > L$, we can find that

$$\begin{aligned} \mathbf{h}_g^H \mathbf{h}_i &= 0, i \geq g + L, \\ \mathbf{h}_i^H \mathbf{h}_g &= 0, i \leq g - L, \end{aligned} \tag{16}$$

and thus, $\mathbf{J}_{k-1}$ has the following form

$$\mathbf{J}_{k-1} = \begin{bmatrix} \mathbf{h}_k^H \mathbf{h}_{k-L_b} & \cdots & \cdots & \mathbf{h}_k^H \mathbf{h}_{k-1} \\ 0 & \mathbf{h}_{k+1}^H \mathbf{h}_{k-L_b+1} & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \mathbf{h}_{k+L_b-1}^H \mathbf{h}_{k-1} \\ 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix},$$

which has size $L_f \times L_b$. We can observe that there are only $\sum_{i=1}^{L_b} i = \frac{1+L_b}{2} L_b$ non-zero elements in $\mathbf{J}_{k-1}$ so that the reconstruction of the detected terms $\sum_{i=k-L_b}^{k-1} \mathbf{j}_i s_i$ can be further simplified. Similarly, in (15), the calculation of $\mathbf{\Lambda}_k$ and $\mathbf{j}_k$ can be simplified as well.

Now, we discuss the complexity of the A-ML-DFBE, linear-MMSE [4], MMSE-DFE [4], and BAD [10] detectors in terms of the number of additions and multiplications. The resulting values are given in Table II, obtained by inspection of the relevant algorithms in Table I, [4], and [10]. Details of the computation of complexity, for example the matrix inversion, can be found in [12]. The computational complexity of the A-ML-DFBE algorithm is a function of the frame length (N), the impulse response length (L), and the length of the forward filter (L_f), which is obtained on the basis of Table. I. From the table, we observe that A-ML-DFBE has the same order of complexity as the linear-MMSE and MMSE-DFE. But A-ML-DFBE is less complex than MMSE-DFE since the A-ML-DFBE requires smaller L_f value, and it does not require to build up the backward filter. In comparison to linear-MMSE, the A-ML-DFBE needs relatively even shorter forward filter and thus has lower complexity. The relation between the filter length and the performance can be clearly observed in the simulation results section. BAD requires complexity at least double of MMSE-DFE. Note that with regard to computational complexity, we focus on time-domain implementation even though a low-complexity frequency-domain implementation is also possible by making use of the block-circulant structure that can be created by the guard interval. In addition, note that the matrix inversion lemma can be used to reduce the complexity from cubic to quadratic order, but it does not affect the above conclusions.

IV. Performance Analysis

A. Analytical SER and BER Derivations

In this subsection, we analyze the symbol error rate (SER) as well as the bit error rate (BER) performance of the A-ML-DFBE. Note that the tail detection only contains the operation of very few symbols, and thus, the performance is dominated by Step 2 of the A-ML-DFBE process in Table I, which will now be analyzed. We assume that all the decisions are accurate for analysis, which is a normal assumption in decision feedback theory [4]. In (12), which contains correlated noise, $\boldsymbol{\eta}_k$, the pre-whitening filter, $\mathbf{\Psi}_k = \mathbf{\Lambda}_k^{-\frac{1}{2}}$, can be applied to make the variance of the noise uncorrelated

$$\mathbf{\Psi}_k \widetilde{\mathbf{y}}_k = \mathbf{\Psi}_k \mathbf{j}_k s_k + \mathbf{\Psi}_k \boldsymbol{\eta}_k, \tag{17}$$

where $\mathbf{\Psi}_k \boldsymbol{\eta}_k$ with size $L_f \times 1$ has a Gaussian distribution with zero mean and all components have unit variance.

Since the noise now has become white Gaussian, the matched filter, $(\mathbf{\Psi}_k \mathbf{j}_k)^H$, can be employed and we have the following received signal equation in scalar form

$$y_k^{'} = \xi_k s_k + \upsilon_k, \tag{18}$$

where $y_k^{'} = (\mathbf{\Psi}_k \mathbf{j}_k)^H \mathbf{\Psi}_k \widetilde{\mathbf{y}}_k$, $\xi_k = \|\mathbf{\Psi}_k \mathbf{j}_k\|^2$, and $\upsilon_k = (\mathbf{\Psi}_k \mathbf{j}_k)^H \mathbf{\Psi}_k \boldsymbol{\eta}_k$, which is a scalar with zero mean and variance $\|\mathbf{\Psi}_k \mathbf{j}_k\|^2$. The SER for M-PSK constellation is given by [13]

$$\mathsf{SER}_M^k = \frac{1}{\pi} \int_0^{\frac{(M-1)\pi}{M}} \mathsf{exp}\left(-\frac{g_{\mathsf{psk}} \gamma_k}{\sin^2 \theta}\right) d\theta, \tag{19}$$

where $g_{\mathsf{psk}} \triangleq \sin^2 \frac{\pi}{M}$, $\gamma_k \triangleq \frac{|\xi_k s_k|^2}{\mathsf{var}(\upsilon_k)} = \frac{\xi_k^2}{(\mathbf{\Psi}_k \mathbf{j}_k)^H (\mathbf{\Psi}_k \mathbf{j}_k)} = \|\mathbf{\Psi}_k \mathbf{j}_k\|^2$, and M denotes the constellation size. The average BER for M-PSK can be written as:

$$\mathsf{BER}_M = \frac{1}{N - L_f + 2} \sum_{k=1}^{N-L_f+2} \mathsf{BER}_M^k, \tag{20}$$

where $\mathsf{BER}_M^k \approx \frac{1}{\log_2 M} \mathsf{SER}_M^k$ [1] for high SNR and Gray mapping. Since the tail is normally short, which has length $L_f - 2$, in comparison to the whole frame length N, hence its effects can be neglected. Note that in time-invariant channel, $\mathsf{SER}_M^1 = \mathsf{SER}_M^2 = \cdots = \mathsf{SER}_M^{N-L_f+2}$ due to the property of $\mathbf{J}$ ($\gamma_1 = \gamma_2 = \cdots = \gamma_{N-L_f+2}$) by assuming perfect decision feedback at high SNR.

B. Multipath Diversity Analysis

Next, we analyze further the behavior of the proposed A-ML-DFBE at high SNR. Assuming perfect channel estimation at the receiver, and taking (19) as an example, it can be upper bounded by [1]

$$\mathsf{SER}_M^k \leq \frac{1}{2} \mathsf{exp}\left(-\frac{g_{\mathsf{psk}}}{\sin^2 \theta} \gamma_k\right) \approx \frac{1}{2} \mathsf{exp}\left(-\frac{g_{\mathsf{psk}}}{\sigma^2 L_f \sin^2 \theta} \sum_{i=0}^{\mathsf{min}(L_f-1, L-1)} |h_i(t)|^2\right), \tag{21}$$

where $\gamma_k \approx \frac{1}{\sigma^2 L_f} \sum_{i=0}^{\mathsf{min}(L_f-1, L-1)} |h_i(t)|^2$ at high SNR (Refer to Appendix I for the derivation). In order to obtain good performance in terms of multipath combining and inter-symbol interference suppression, we should choose $L_f \geq L$. Then, by averaging (21) over the Rayleigh PDF [14], equation (21) becomes

$$\overline{\mathsf{SER}_M^k} \leqslant \frac{1}{2} \left(\frac{g_{\mathsf{psk}} \mathsf{SNR}}{L \cdot L_f \sin^2 \theta}\right)^{-L}, \tag{22}$$

which indicates that the A-ML-DFBE achieve the maximum multipath diversity order L.

C. Analysis of the Length of the Forward Filter (Sliding Window) and Backward Filter

It has been shown that the forward filter length, L_f, is a very important parameter in the proposed A-ML-DFBE. In this subsection, we discuss the behaviors of L_f: 1) Increasing the value of L_f can improve the robustness of (15) due to the following reasons: Firstly, as shown in (5), larger value of L_f can incorporate more received signals as well as channel information in the forward filter; Secondly, indicated by (13), increasing L_f can make the Gaussian assumption more accurate; 2) While the performance can be enhanced, as shown in Table II, the complexity will correspondingly go up. Hence, for A-ML-DFBE, a complexity and performance tradeoff can be realized by adjusting L_f; 3) Performance gets converged by increasing the value of L_f as the Gaussian assumption becomes accurate enough. This implies that moderate length of the forward filter can deliver good performance;

4) Given by Subsection-IV-B, L_f should be equal or larger than L for maximum diversity order; 5) The length of the backward, L_b, always equals $L-1$ due to the special structure of $\mathbf{H}$.

D. Analysis of the Matched Filter in (3)

Note that the matched filter in (3) can obtain some additional information from the received signals outside the slicing window. Recalling (4)–(9), $\mathbf{y}_k$ can be written as

$$\mathbf{y}_k = [\mathbf{h}_k^H \mathbf{r}, \ldots, \mathbf{h}_{k+L_f-1}^H \mathbf{r}]^T. \tag{23}$$

Although some information is lost after horizonal and vertical slicing, some gains can be still realized by considering the whole received signal, $\mathbf{r}$.

Supposing the matched filter is removed, the detection procedures in Table I can be used but it will lead to performance degradation since only the received signals inside the sliding window will be considered, where $\mathbf{y}_k = [r_k, \ldots, r_{k+L_f-1}]^T$. As a result, the length of the forward filter has to be increased to make up the performance loss caused by the slicing processes in order to obtain the same performance. Note also if the length of the forward filter is equal to N, the A-ML-DFBE directly enters the 'Tail Detection' step (Step **3**) in Table I, which will make no difference in performance whether or not the matched filter is used since there is no slicing operations at all. However, the value of L_f is normally much less than N. Theoretically, using the same methods as shown in Appendix I, it is easy to obtain the SNR for the A-ML-DFBE when the matched filter is removed. Due to the space limitation, we drop the detailed derivation part. But we can conclude that the performance of A-ML-DFBE can be upper-bounded by the same equalize without using the matched filter.

V. Simulation Results

In all simulations, BPSK constellation is used to generate a rate 1bps/Hz transmission. We plot the BER versus the signal-to-noise ratio (SNR). For analytical results, we assume perfect decision feedback, but for simulated results we use the feedback decisions. The performance is determined over doubly selective Rayleigh fading channels. The impulse response length is $L = 5$, and, thus, the length of the backward filter of the A-ML-DFBE can be fixed as $L_b = L - 1 = 4$. Jakes' Model is applied to construct time-selective Rayleigh fading channel for each subpath. The carrier frequency $f_c = 2$ GHz and the symbol period $T_s = 128/c$, where c is the speed of light. The simulation results are plotted with two speeds: $v = 5\ vkm/h$ and $150\ km/h$ (corresponding to $f_d T_s = 0.0001$ and 0.0093, where Doppler frequency $f_d = v f_c/c$). The frame length N is 128.

In Fig. 1 and Fig. 2, we examine the analytical BER performance obtained in (20) assuming that the channel estimation is perfect. The simulations are plotted with the vehicle speed: $v = 5km/h$. In Fig. 1, we compare the analytical BER with the simulated BER. It can be observed that the analytical

BER is close and asymptotically converges to the simulated curves at high SNR. In Fig. 2, the analytical BER for A-ML-DFBE is plotted employing different forward filter lengths. As discussed earlier, the length of the forward filter, L_f, should be at least equal to L in order to realize good performance. From Fig. 2, we can see that the proposed A-ML-DFBE with $L_f = 5$ provides much better performance than that with $L_f = 3$, and as the value of L_f increases, the performance begins to converge. It can be also seen that for A-ML-DFBE, $L_f = 10$ (two times L) is enough to obtain good BER performance.

In Fig. 3, simulation results for the A-ML-DFBE detector are illustrated in comparison with conventional linear MMSE, MMSE-DFE, BAD, and MLSE decoders. The simulations are plotted with the vehicle speed: $v = 5\ km/h$. Least square (LS) channel estimation [4] is used. From Fig. 3, it can be observed that at BER=10^{-3}, the performance of A-ML-DFBE with $L_f = 5$ is far better than the linear MMSE and the MMSE-DFE equalizers. There is only 2 dB loss compared to the MLSE decoder at BER=10^{-5}. At $L_f = 10$, there is about 0.8 dB loss compared to MLSE. Almost no difference can be observed for A-ML-DFBE when L_f is increased to 15 since $L_f = 10$ is sufficient to make the performance converge. Note that when $L_f = 15$, A-ML-DFBE gives almost the same performance as $L_f = 10$, which demonstrates that only a small value of L_f is required to achieve good performance. We can also see that A-ML-DFBE with $L_f = 5$ can provide much better performance than BAD with $L_f = 15$. Note that our A-ML-DFBE has lower complexity than MMSE-DFE, and thus, lower than BAD. Clearly, from Fig. 1 to Fig. 3,we can see that there exists a complexity and performance tradeoff in terms of L_f. Performance can be improved by increasing the length of the forward filter (slicing window). In addition, performance convergence can be also observed, which indicates that limited value of L_f is enough to deliver most of the performance gain.

In Fig. 4, simulation comparisons are made for A-ML-DFBE without using the matched filter. Perfect channel estimation is assumed. Vehicle speed, $v = 5km/h$, is adopted. We choose different L_f values for the no matched filter case: 5, 10, and 15 and L_b remains the same: 4. It is shown that at $L_f = 5$, the performance without the matched filter is worse than with it. We can also observe the significant performance loss due to the small value of L_f. It is shown that L_f must be 15 for the system with no matched filter to provide the same performance as the matched filter system with $L_f = 10$. Hence, from the simulation results we can see that the matched filter is very important for system performance. Note that as discussed in the complexity analysis part, Subsection III-B, the forward and backward filter taps are actually fixed and can be obtained before the A-ML-DFBE detection. The complexity increase by the use of the matched filter is much more worthwhile than to increase the length of the forward filter without using the matched filter.

In Fig. 5, simulation results for the A-ML-DFE detector are illustrated in comparison with conventional linear MMSE, MMSE-DFE, BAD, and MLSE decoders using LS channel estimation and the

vehicle speed is $v = 150km/h$. Here, we choose different values for L_f for A-ML-DFBE. From the simulation results, we can still observe that the performance of A-ML-DFBE converged at $L_f = 10$, and no gain can be obtained at $L_f = 15$. Due to the time-variant effects, the performance is degraded compared to the results in Fig. 3. We can see about 1 dB loss between MLSE and A-ML-DFBE with $L_f = 10$ when BER=10^{-5}. However, the proposed equalizer can still substantially outperform Linear MMSE and MMSE-DFE in all SNR regime. Around 8 dB performance gain can be obtained by the proposed scheme with $L_f = 5$ compared to the BAD at BER=10^{-3}.

VI. Conclusions

In this paper, we have proposed a simple approximate ML decision feedback equalizer for doubly selective fading environment. From the analytical and simulation results, we conclude that the A-ML-DFBE significantly outperforms the linear MMSE, MMSE-DFE, and BAD detectors, and provides performance very close to MLSE. We have shown that when L_f is large enough, further increases in L_f do not improve performance much. This implies that the proposed equalizer is quite robust against ISI. A tradeoff in terms of the complexity and the performance can be achieved by adjusting the value of L_f. Computational complexity comparison has demonstrated that the A-ML-DFBE requires fewer additions and multiplications than MMSE based schemes. In addition, the implementation of the matched filter is very important and the A-ML-DFBE obtains maximum diversity order when $L_f \geq L$.

Due to the DFE processing, parallel computing is difficult to achieve for the proposed equalizer. However, by adjusting the size of the data block or the filters (back and forward), or both, the latency can be reduced. The proposed equalizer can be easily used for radar communication systems as it is suitable to solve time-domain equalization problems. In current wireless systems like UMTS, HSDPA or HSUPA, the A-ML-DFBE can be used to recover signals similar to MMSE or MMSE-DFE. For LTE or LTE advance, the proposed algorithm can be extended to realize frequency-domain equalizations.

Appendix A
Derivation of Closed-Form Expression of γ_k at High SNR

Now, the closed-form expression of γ_k at high SNR is derived in terms of L_f and L. From Subsection IV-A, γ_k can be written as

$$\gamma_k = \|\mathbf{\Psi}_k \mathbf{j}_k\|^2 = \mathbf{j}_k^H \mathbf{\Lambda}_k^{-1} \mathbf{j}_k = \mathbf{j}_k^H \left(\sigma^2 \mathbf{X} + \mathbf{Y}\mathbf{Y}^H\right)^{-1} \mathbf{j}_k, \tag{24}$$

where for convenience $\mathbf{Y} \triangleq \mathbf{J}_{k+1}$ has size $L_f \times L_f - 1$, and $\mathbf{X} \triangleq \mathbf{J}_k^{'}$. By using the Kailath Variant $(\mathbf{A} + \mathbf{BC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}$ [15], the inversion term on the right side of (24) can be further written as

$$\gamma_k = \sigma^{-2}\mathbf{X}^{-1} - \sigma^{-2}\mathbf{X}^{-1}\mathbf{Y}\left(\sigma^2\mathbf{I}_{L_f-1} + \mathbf{Y}^H\mathbf{X}^{-1}\mathbf{Y}\right)^{-1}\mathbf{Y}^H\mathbf{X}^{-1}. \tag{25}$$

At high SNR, as $\sigma^2 \to 0^+$, the effect of $\sigma^2\mathbf{I}_{L_f-1}$ is comparatively small, which can be ignored from an asymptotic point of view. Hence, we have the following approximation for the second term in (25)

$$\sigma^{-2}\mathbf{X}^{-1}\mathbf{Y}(\mathbf{Y}^H\mathbf{X}^{-1}\mathbf{Y})^{-1}\mathbf{Y}^H\mathbf{X}^{-1} = \sigma^{-2}\mathbf{X}^{-\frac{1}{2}}\mathbf{Z}\left(\mathbf{Z}^H\mathbf{Z}\right)^{-1}\mathbf{Z}^H\mathbf{X}^{-\frac{1}{2}}. \tag{26}$$

where $\mathbf{Z} \triangleq \mathbf{X}^{-\frac{1}{2}}\mathbf{Y}$ with size $L_f \times (L_f - 1)$, and $(\cdot)^{-\frac{1}{2}}$ represents the unique positive definite Hermitian root [15].

Let $\mathbf{Z}^+$ be the Moore-Penrose inverse of matrix $\mathbf{Z}$, and $\mathbf{Z}^+ = (\mathbf{Z}^H\mathbf{Z})^{-1}\mathbf{Z}^H$ of size $(L_f - 1) \times L_f$. Note that $\mathtt{rank}(\mathbf{Z}\mathbf{Z}^+) = \mathtt{rank}(\mathbf{Y}) = L_f - 1$ and $\mathbf{Z}\mathbf{Z}^+$ has size $L_f \times L_f$. By eigenvalue decomposition, we can get $\mathbf{Z}\mathbf{Z}^+ = \mathbf{U}\mathbf{\Pi}\mathbf{U}^H$ where $\mathbf{U}$ is the unitary eigenvector matrix and $\mathbf{\Pi} \triangleq \mathtt{diag}\{\lambda_1, \ldots, \lambda_{L_f-1}, 0\}$. From the definition of $\mathbf{Z}\mathbf{Z}^+$, we have $(\mathbf{Z}\mathbf{Z}^+)^2 = \mathbf{Z}\mathbf{Z}^+$. Therefore, $\mathbf{Z}\mathbf{Z}^+$ is idempotent [15], and any idempotent matrix has eigenvalue 1 or 0, and thus $\mathbf{\Pi} = \mathtt{diag}\{1, \ldots, 1, 0\}$. We can then get

$$\mathbf{j}_k^H\mathbf{X}^{-\frac{1}{2}}\mathbf{Z}\left(\mathbf{Z}^H\mathbf{Z}\right)^{-1}\mathbf{Z}^H\mathbf{X}^{-\frac{1}{2}}\mathbf{j}_k = \mathbf{j}_k^H\mathbf{X}^{-\frac{1}{2}}\mathbf{U}\mathbf{\Pi}\mathbf{U}^H\mathbf{X}^{-\frac{1}{2}}\mathbf{j}_k \approx \frac{L_f - 1}{L_f}\mathbf{j}_k^H\mathbf{X}^{-1}\mathbf{j}_k. \tag{27}$$

From (24), (25), (26), and (27), at high SNR, we can obtain

$$\gamma_k \approx \frac{1}{\sigma^2 L_f}\mathbf{j}_k^H\mathbf{X}^{-1}\mathbf{j}_k, \tag{28}$$

From (5), we can get

$$\mathbf{j}_k^H\mathbf{X}^{-1}\mathbf{j}_k = \mathbf{h}_k^H\mathbf{H}_k(\mathbf{H}_k^H\mathbf{H}_k)^{-1}\mathbf{H}_k^H\mathbf{h}_k, \tag{29}$$

where $\mathbf{j}_k = \mathbf{H}_k^H\mathbf{h}_k$ and $\mathbf{X} = \mathbf{H}_k^H\mathbf{H}_k$ has size $L_f \times L_f$ and rank L_f. Since $\mathbf{H}_k(\mathbf{H}_k^H\mathbf{H}_k)^{-1}\mathbf{H}_k^H$ has the same structure as $\mathbf{Z}\left(\mathbf{Z}^H\mathbf{Z}\right)^{-1}\mathbf{Z}^H$ in (27), we can get the corresponding eigenvalues as

$$\mathtt{EIG}\left(\mathbf{H}_k(\mathbf{H}_k^H\mathbf{H}_k)^{-1}\mathbf{H}_k^H\right) = \mathtt{diag}\{\underbrace{0, \ldots, 0}_{k-1}, \underbrace{1, \ldots, 1}_{L_f}, \underbrace{0, \ldots, 0}_{N+L-L_f-k}\}. \tag{30}$$

Finally, combining (28) and (29), at high SNR as $\sigma^2 \to 0^+$, finally we have

$$\gamma_k \approx \frac{1}{\sigma^2 L_f} \sum_{i=0}^{\mathtt{min}(L_f-1, L-1)} |h_i(t)|^2. \tag{31}$$

REFERENCES

[1] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, USA, 2001.

[2] G. D. Forney, "Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," *IEEE Trans. Info. Theory*, vol. IT-18, no. 3, pp. 363-378, May 1972.

[3] G. K. Kaleh, "Channel Equalization for Block Transmission Systems," *IEEE J. Select. Areas Commun.*, vol. 13, no. 1, pp. 110–121, Jan. 1995.

[4] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Prentice-Hall, Inc; Upper Saddle River, NJ, USA, 1996.

[5] G. Leus and M. Moonen, *Equalization Techniques for Fading Channels*. Chapter in Handbook on Signal Processing for Communications (M. Ibnkahla ed.), CRC Press, 2004

[6] L. Song and A. G. Burr, "Suceesive Interference Cancellation for Space-Time Block Codes over Time-Selective Fading Channels," *IEEE Commun. Letters*, vol. 10, no. 12, pp. 837–839, Dec. 2006.

[7] Y. Jia, C. Andrieu, R. J. Piechocki, and M. Sandell, "Gaussian Approximation Based Mixture Reduction for Near Optimum Detection in MIMO Systems," *IEEE Commun. Letters*, vol. 9, no. 11, pp. 997–999, Nov. 2005.

[8] J. Luo, K. R. Pattipati, P. K. Willett, and F. Hasegawa, "Near-Optimal Multiuser Detection in Synchronous CDMA Using Probabilistic Data Association," *IEEE Commun. Letters*, vol. 5, no. 9, pp. 361–363, Sep. 2001.

[9] S. Liu and Z. Tian, "Near-Optimum Soft Decision Equalization for Frequency Selective MIMO Channels," *IEEE Trans. Signal Processing*, vol. 5, no. 3, pp. 721–733, Mar. 2004.

[10] J. K. Nelson, A. C. Singer, and U. Madhow, "BAD: Bidirectional Arbitrated Decision-Feedback Equalization," *IEEE Trans. Commun.*, vol. 53, no. 2, pp. 214–218, Feb. 2005.

[11] A. M. Chan and G. W. Wornell, "A Class of Block-Iterative Equalizers for Intersymbol Interference Channels: Fixed Channel Results," *IEEE Trans. on Commun.*, vol. 49, no. 11, pp. 1966–1976, Nov. 2001

[12] G. H. Golub and C. D. Loan, *Matrix Computations*, 3rd ed. Johns Hopkins University Press, Baltimore, MD, 1996.

[13] M. K. Simon and M.-S. Alouini, *Digital Communications over Fading Channels: A Unified Approach to Performance Analysis*. Wiley Series in Telecommunications and Signal Processing, USA, 2001.

[14] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-Time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[15] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, United Kingdom, 1991, reprinted 1999.

TABLE I
APPROXIMATE MAXIMUM LIKELIHOOD DECISION FEEDBACK BLOCK EQUALIZATION ALGORITHM.

Step 1: Initialization of A-ML-DFBE

1. Set L_f ($L_f \geq L$).
2. Fix $L_b = L - 1$.

Step 2: A-ML-DFBE Detection

For $k = 1 : N - L_f$

1. `Forward Process`: Apply the temporal sub matched filter according to (4).
2. `Decision Feedback Process`: Remove the effects reconstructed by the past decisions using (10).
3. `Approximate ML`: Recover the transmitted signals with (15).

end;

Step 3: Tail Detection

1. Fix the sliding window:
   $\mathbf{y}_{N-L_f+1} = [y_{N-L_f+1}, \ldots, y_N]^T$.
2. Signal Recovery:

For $k = N - L_f + 1 : N$

1. `Decision Feedback Process`: Remove the effects of the past decisions using (10).
2. `Approximate ML`: Recover the transmitted signals with (15).

end;

TABLE II

COMPUTATIONAL COMPLEXITY OF VARIOUS SCHEMES FOR ONE SLIDING WINDOW WITH LENGTH N; L IS THE NUMBER OF PATHS; BPSK CONSTELLATIONS; L_f AND L_b STAND FOR THE LENGTH OF THE FORWARD AND BACKWARD FILTERS, RESPECTIVELY.

Detector	Additions	Multiplications
A-ML-DFBE	$N[8L_f^3+34L_f^2+(6L+7)L_f+(3L-1)]$	$N[(2L_f^3+42L_f^2)-(12L+19)L_f+18]$
Linear-MMSE	$N[8L_f^3+30L_f^2+2(3L+2)L_f]$	$N[2L_f^3+42L_f^2-(12L-17)L_f-(6L-1)]$
MMSE-DFE	$N[8(L_f^3+L_b^3)+42(L_f^2+L_b^2)+2(3L+2)(L_f+L_b)]$	$N[2(L_f^3+L_b^3)+42(L_f^2+L_b^2)+(12L-11)(L_f+L_b)+6]$
BAD	$N[16(L_f^3+L_b^3)+84(L_f^2+L_b^2)+4(3L+2)(L_f+L_b)]$	$N[4(L_f^3+L_b^3)+84(L_f^2+L_b^2)+2(12L-11)(L_f+L_b)+12]$

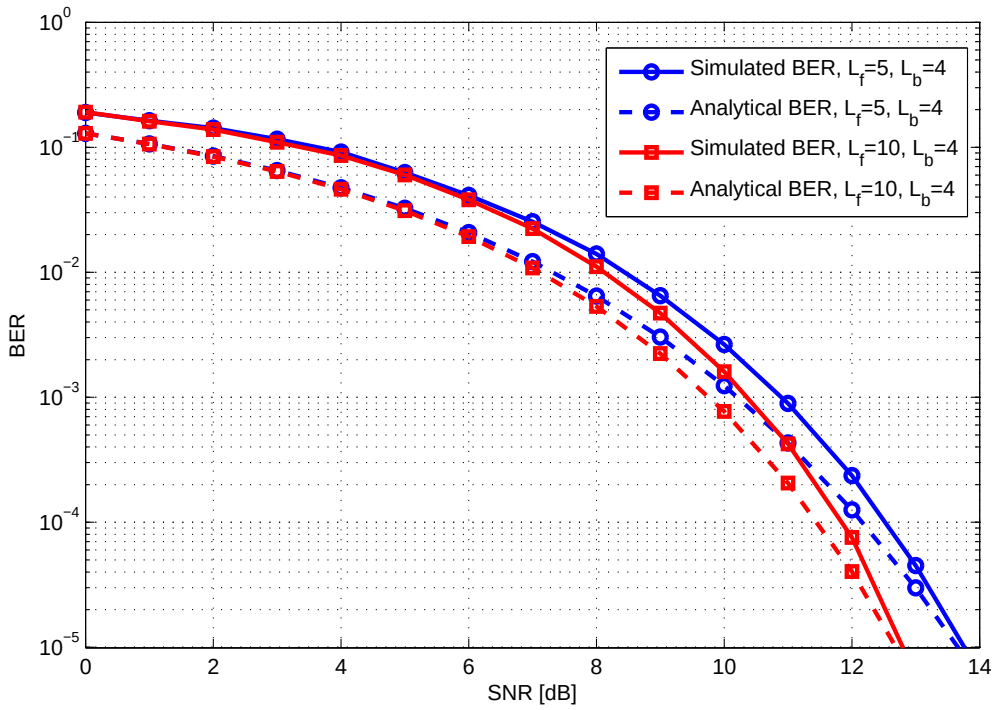

Fig. 1. Analytical BER performance of the A-ML-DFBE over a doubly selective fading channel with perfect channel estimation ($L = 5$, $f_d T_s = 0.0001$) and simulated BER.

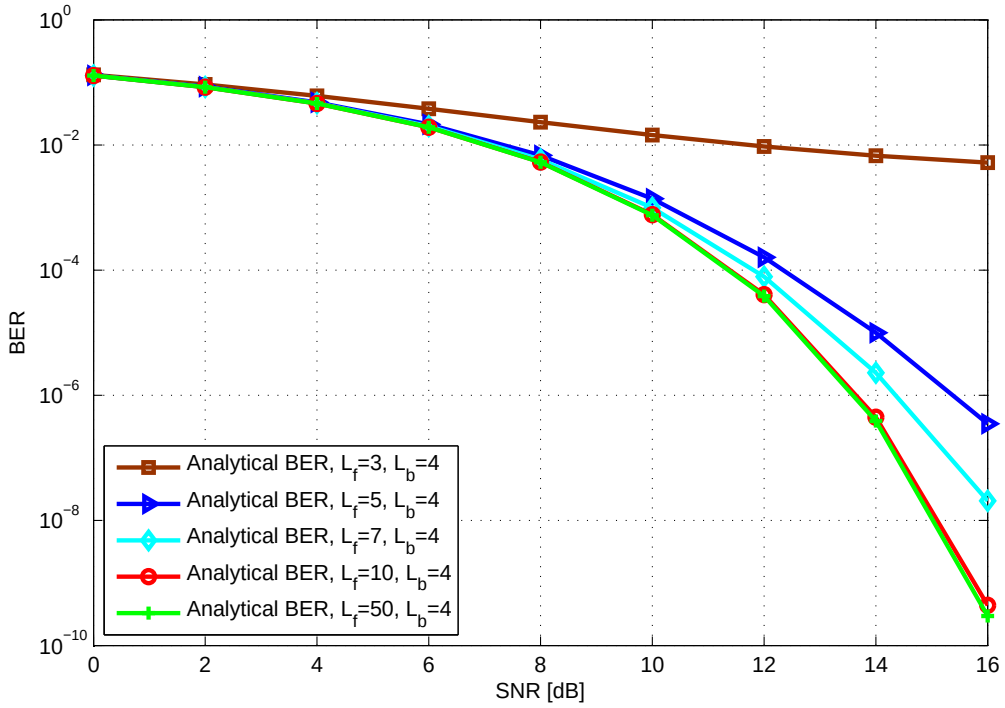

Fig. 2. Analytical BER performance of the A-ML-DFBE with various forward filter length over a doubly selective fading channel with perfect channel estimation ($L = 5$, $f_d T_s = 0.0001$).

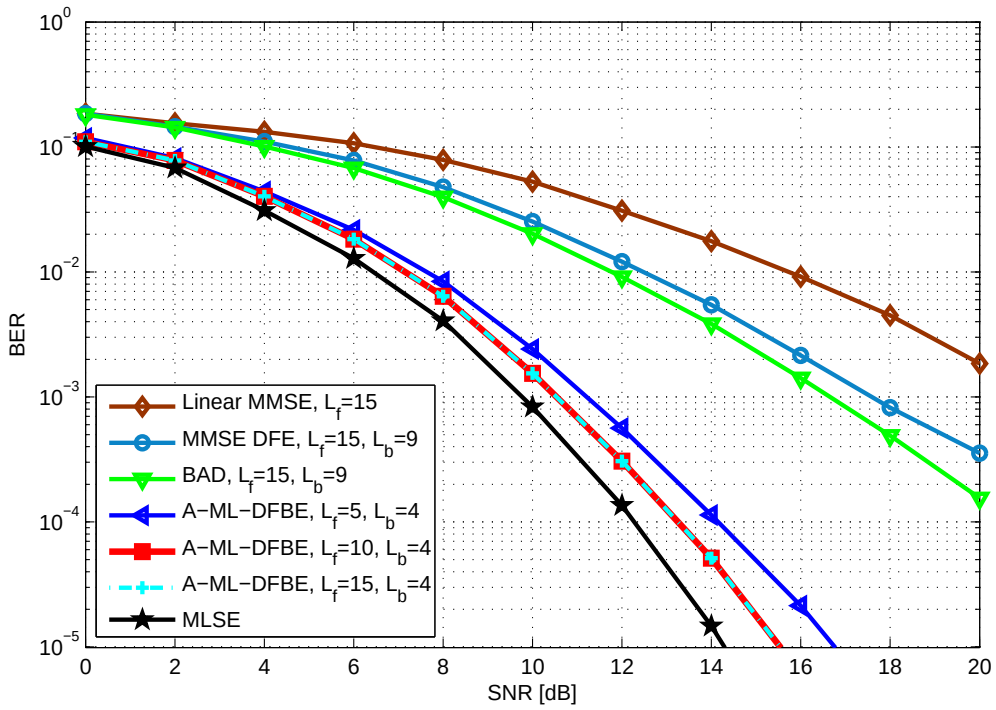

Fig. 3. Simulated BER performance of the A-ML-DFBE over a doubly selective fading channel with LS channel estimation ($L = 5$, $f_d T_s = 0.0001$). Shown for comparisons are the Linear MMSE, MMSE-DFE, BAD, and MLSE.

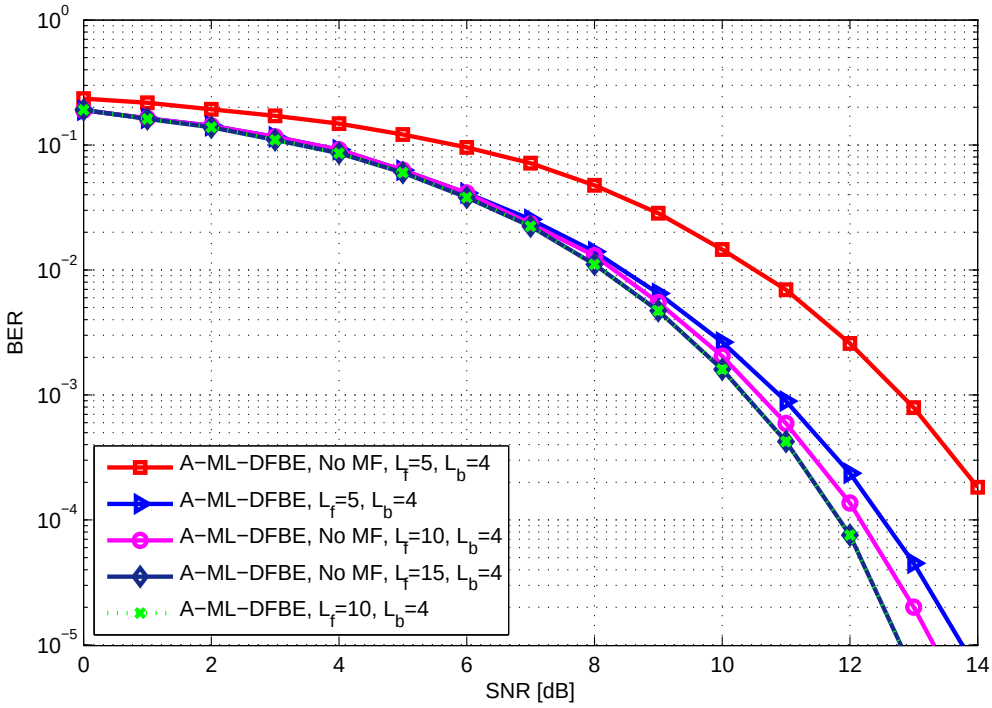

Fig. 4. Simulated BER performance of the A-ML-DFBE over a doubly selective fading channel with perfect channel estimation ($L = 5$, $f_d T_s = 0.0001$). Shown for comparisons are A-ML-DFBE with and without matched filter (MF).

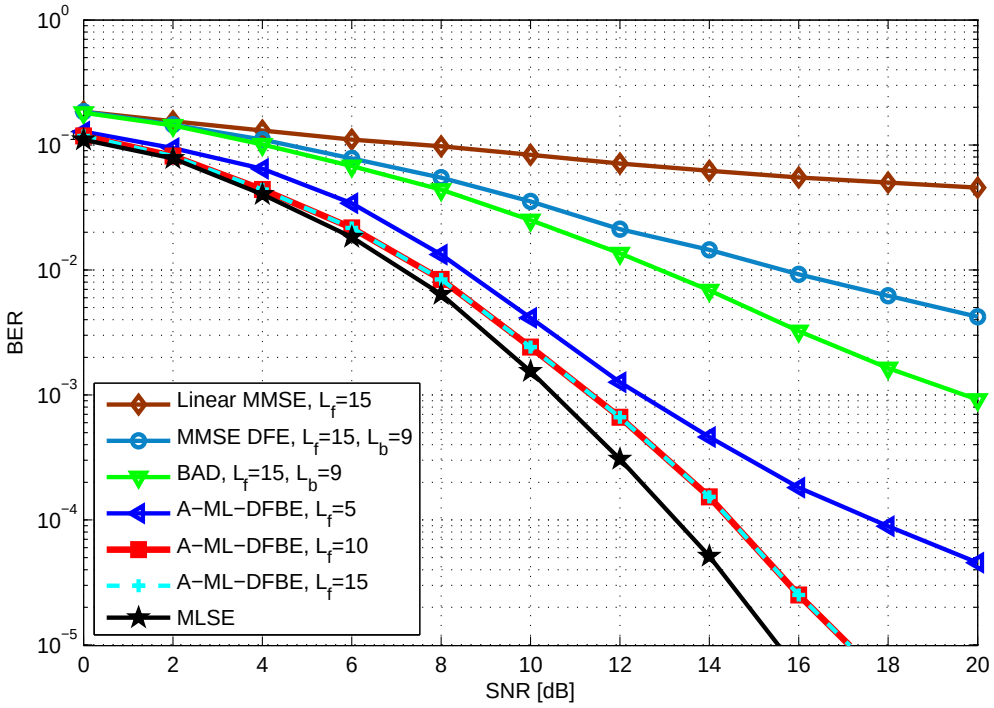

Fig. 5. Performance of the A-ML-DFBE over a doubly selective fading channel with LS channel estimation ($L = 5$, $f_d T_s = 0.0093$). Shown for comparisons are the Linear MMSE, MMSE-DFE, BAD, and MLSE.